

# Saturation models of HERA DIS data and inclusive hadron distributions in p+p collisions at the LHC

Prithwish Tribedy<sup>1</sup> and Raju Venugopalan<sup>2</sup>

<sup>1</sup>*Variable Energy Cyclotron Centre, 1/AF Bidhan Nagar, Kolkata-700064, India*

<sup>2</sup>*Physics Dept., Bldg. 510A, Brookhaven National Laboratory, Upton, NY 11973, USA*

Unintegrated gluon distributions sensitive to the transverse spatial distribution of gluons in the proton are extracted from data on exclusive and diffractive final states at HERA in the dipole approach. These unintegrated gluon distributions can be used to compute inclusive hadron production in p+p collisions at the LHC. In this paper, we consider a number of saturation models with differing dynamical assumptions that give good fits to the available HERA data. We apply these models to study the rapidity and transverse momentum dependence of the LHC data up to  $\sqrt{s} = 7$  TeV. We examine the sensitivity of these results to parameters that are not constrained by the HERA data and comment on similarities and differences with previous work. We compute the n-particle inclusive multiplicity distribution and show that the LHC p+p results are in agreement with predictions for multi-particle production in the Color Glass Condensate approach. This result has significant ramifications for the interpretation of multi-particle correlations in high multiplicity events at the LHC.

## I. INTRODUCTION

Colliders at ultra-high energies have made feasible the study of the collective many-body properties of QCD in the limit of fixed  $Q^2 \gg \Lambda_{\text{QCD}}^2$  and very small  $x$ . An important property of QCD in these “Regge asymptotics” is the phenomenon of gluon saturation, which states that a probe with transverse resolution  $1/Q^2$  interacts with a target with probability of order unity when  $Q^2 \leq Q_S^2(x)$ , where  $Q_S^2(x)$  is a dynamically generated semi-hard scale in the target hadron or nucleus [1]. In the framework where the parton model is manifest, the gluon saturation phenomenon corresponds to the requirement that transverse momentum modes with  $k_\perp \leq Q_S$  are maximally occupied with an occupation number  $1/\alpha_s$ . The large occupation number means that these modes can be treated classically [2]; an additional separation of scale between fast and slow modes due to time dilation allows for a renormalization group treatment of these high occupation number states as they evolve with energy [3]. This Color Glass Condensate (CGC) [4] description of saturated gluons is universal to hadrons and nuclei and can be tested both in deeply inelastic scattering experiments off hadrons and nuclei as well as in hadron/nuclear collisions.

The CGC approach, when applied to Deeply Inelastic Scattering (DIS), results at leading order in  $\alpha_s$  [5] in the dipole picture where the inclusive virtual photon hadron cross section is expressed as [6]

$$\sigma_{L,T}^{\gamma^*p} = \int d^2\mathbf{r}_\perp \int_0^1 dz \left| \Psi_{L,T}^* \right|^2 \int d^2\mathbf{b}_\perp \frac{d\sigma_{\text{dip}}^p}{d^2\mathbf{b}_\perp}. \quad (1)$$

Here  $\left| \Psi_{L,T}^*(\mathbf{r}_\perp, z, Q) \right|^2$  represents the probability for a virtual photon to produce a quark–anti-quark pair of size  $r = |\mathbf{r}_\perp|$  and  $\frac{d\sigma_{\text{dip}}^p}{d^2\mathbf{b}_\perp}(\mathbf{r}_\perp, x, \mathbf{b}_\perp)$  denotes the *dipole cross section* for this pair to scatter off the target at an impact parameter  $\mathbf{b}_\perp$ . The former is well known from QED, while the latter represents the dynamics of QCD scattering at small  $x$ . It was shown some time ago that a model known as the Golec-Biernat–Wusthoff (GBW) model [7] that implements saturation in the dipole cross-section through the parametrization  $\frac{d\sigma_{\text{dip}}^p}{d^2\mathbf{b}_\perp} = 2(1 - e^{-r^2 Q_{s,p}^2(x)/4})$ , where  $Q_{s,p}^2(x) = (x_0/x)^\lambda \text{ GeV}^2$ , gives a good qualitative fit to the HERA inclusive and diffractive cross section data for  $x_0 = 3 \cdot 10^{-4}$  and  $\lambda = 0.288$ . This work explained very simply key features of the HERA data and was very suggestive of the possible role of a semi-hard saturation scale in the hadron. This work was further refined in more sophisticated models that are somewhat better motivated and treat the impact parameter dependence of the dipole cross-section more accurately. As we shall discuss further in the next section, these models give excellent fits to small  $x$  inclusive, diffractive and exclusive HERA data. The common ingredient in these combined fits is the dipole cross-section.

The dipole cross-section, to leading logarithmic accuracy, is a universal quantity which can be applied to compute inclusive quantities in hadron-hadron collisions. It is defined in terms of the real part of the forward scattering amplitude  $\mathcal{N}(\mathbf{r}_\perp, x, \mathbf{b}_\perp)$  as

$$\frac{d\sigma_{\text{dip}}^p}{d^2\mathbf{b}_\perp}(\mathbf{r}_\perp, x, \mathbf{b}_\perp) = 2 \mathcal{N}(\mathbf{r}_\perp, x, \mathbf{b}_\perp) \equiv 2 \left( 1 - \frac{1}{N_c} \left\langle \text{tr} \left( \tilde{U}(\mathbf{b}_\perp + \frac{\mathbf{r}_\perp}{2}) \tilde{U}^\dagger(\mathbf{b}_\perp - \frac{\mathbf{r}_\perp}{2}) \right) \right\rangle_x \right), \quad (2)$$

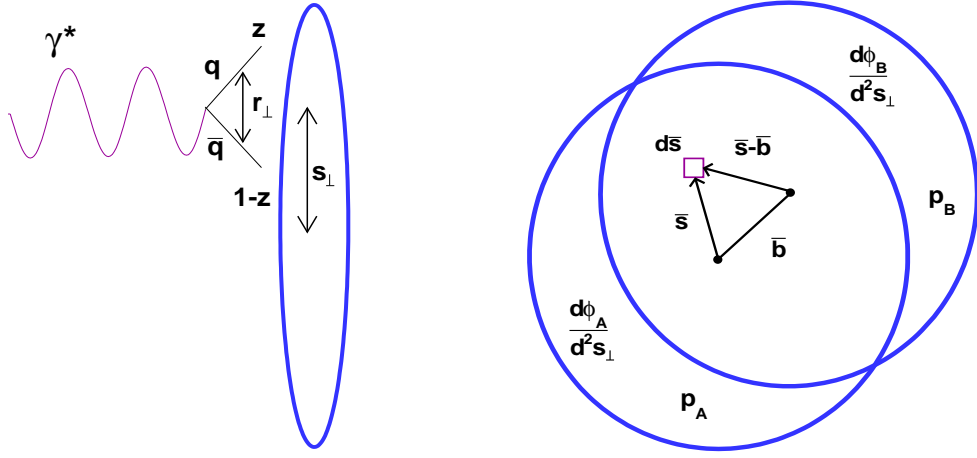


FIG. 1: Left: Dipole cross-section in DIS. Right: Overlap of unintegrated gluon distributions in proton-proton collisions.

where  $\tilde{U}(\mathbf{b}_\perp \pm \frac{\mathbf{r}_\perp}{2})$  is a Wilson line in the fundamental representation representing the interaction between a quark and the color fields of the target. The average  $\langle \dots \rangle_x$  is an average over these color fields; the energy dependence of the correlator as a function of  $x$  (or the rapidity  $Y = \ln(1/x)$ ) is given by the JIMWLK equation [3]. In the large  $N_c$  limit, the equation for the energy evolution of this correlator is the Balitsky-Kovchegov (BK) equation [8]. We note however that neither JIMWLK nor BK is at present equipped to deal well with the impact parameter dependence of the dipole cross-section; the dipole cross-section in this formalism is taken in eq. (2) to be independent of the impact parameter. To address the impact parameter dependence of this equation, one resorts to models which parametrize both saturation effects and the impact parameter dependence.

In hadron-hadron collisions, one can derive at leading order the expression [9]

$$\frac{dN_g(\mathbf{b}_\perp)}{dy d^2\mathbf{p}_\perp} = \frac{16\alpha_S}{\pi C_F} \frac{1}{p_\perp^2} \int \frac{d^2\mathbf{k}_\perp}{(2\pi)^5} \int d^2\mathbf{s}_\perp \frac{d\phi_A(x_1, \mathbf{k}_\perp | \mathbf{s}_\perp)}{d^2\mathbf{s}_\perp} \frac{d\phi_B(x_2, \mathbf{p}_\perp - \mathbf{k}_\perp | \mathbf{s}_\perp - \mathbf{b}_\perp)}{d^2\mathbf{s}_\perp} \quad (3)$$

This equation is a generalization of the well known  $k_\perp$  factorization expression for inclusive gluon production [10] to include the impact parameter dependence of the unintegrated gluon distributions. Here  $C_F = N_c^2 - 1/2N_c$  is the Casimir for the fundamental representation. Using a relation between quark and gluon dipole amplitudes strictly valid in the large  $N_c$  limit, the unintegrated gluon distribution in either of the two protons can be expressed in terms of the corresponding dipole cross-section measured in DIS as [11]

$$\frac{d\phi(x, \mathbf{k}_\perp | \mathbf{s}_\perp)}{d^2\mathbf{s}_\perp} = \frac{\mathbf{k}_\perp^2 N_c}{4\alpha_S} \int_0^\infty d^2\mathbf{r}_\perp e^{i\mathbf{k}_\perp \cdot \mathbf{r}_\perp} \left[ 1 - \frac{1}{2} \frac{d\sigma_{\text{dip}}^p(\mathbf{r}_\perp, x, \mathbf{s}_\perp)}{d^2\mathbf{s}_\perp} \right]^2 \quad (4)$$

Thus the impact parameter dependent dipole cross-section determined from HERA data can be used to compute the single inclusive gluon distribution in proton-proton collisions with no additional parameters. This statement is strictly valid to leading log accuracy for momenta  $k_\perp > Q_{s,p}^2$ . Further, as we shall discuss later, there will be additional parameters that come in when one wants to make contact with the measured hadron spectrum.

This approach was applied most recently to compute the single inclusive hadron spectrum in proton-proton collisions at the LHC by Levin and Rezaeian [12]. The quantitative differences of our study to their work are the following: a) we consider three dipole models that give good fits to HERA data to see whether they give results consistent with the LHC data, b) we study and comment on the dependence of the results on variations of the parameters in the study and c) we convolve the inclusive gluon distribution with fragmentation function instead of using a simple fragmentation prescription as in ref. [12]. We shall also comment on other quantitative differences in our respective treatments at the appropriate points in the text. A qualitative difference of our work relative to that of ref. [12] is that we compute directly the average inclusive multiplicity at a given impact parameter. In computing the minimum bias single inclusive multiplicity distribution, there are similar uncertainties as ref. [12], which can be fixed by normalizing the data to single inclusive data at lower center of mass energies. However, as we shall discuss later, the average multiplicity at a given impact parameter is an essential input in computing the probability distribution as a function of event multiplicity. We shall compute the  $n$ -particle probability distribution and compare our results with the p+p collider data. These results will be important in understanding the role of various sources of fluctuations in the p+p collider data.

## II. PROTON DIPOLE CROSS-SECTION

In this section, we shall discuss some of the dipole model parametrizations that have been compared extensively to the HERA data. We shall consider respectively the IP-Sat model [16], the b-CGC model [13–15] and the NLO-BK model [17]. Our list of dipole parametrizations is by no means exhaustive [18] but is a sample of some of the key approaches where extensive comparisons have been made to the data. We note that dipole models have been applied to understand RHIC data [19], and there have also been preliminary attempts at a combined analysis of RHIC, HERA and fixed target e+A data [20–24]. We also note that there are leading twist models with some similarities to dipole models that have been compared to the HERA data [25].

### A. The IP-Sat Model

The impact parameter dependent dipole saturation model (IP-Sat) [16] is a refinement of the Golec-Biernat–Wusthoff dipole model [7] to give the right perturbative limit when  $\mathbf{r}_\perp \rightarrow 0$  [26]. It is equivalent to the expression derived in the classical effective theory of the CGC, to leading logarithmic accuracy [5]. The proton dipole cross-section in this model is expressed as

$$\frac{d\sigma_{\text{dip}}^p(\mathbf{r}_\perp, x, \mathbf{b}_\perp)}{d^2\mathbf{b}_\perp} = 2 \left[ 1 - \exp \left( -\frac{\pi^2}{2N_c} \mathbf{r}_\perp^2 \alpha_S(\mu^2) xg(x, \mu^2) T_p(\mathbf{b}_\perp) \right) \right]. \quad (5)$$

Here the scale  $\mu^2$  is related to dipole radius  $\mathbf{r}_\perp$  (see fig. 1) as

$$\mu^2 = \frac{4}{\mathbf{r}_\perp^2} + \mu_0^2, \quad (6)$$

where the leading order expression for the running coupling is

$$\alpha_S(\mu^2) = \frac{12\pi}{(33 - 2n_f) \log(\mu^2/\Lambda_{QCD}^2)} \quad (7)$$

with  $n_f=3$ ,  $\Lambda_{QCD}=0.2$  GeV. The model includes saturation as eikonalized power corrections to the DGLAP leading twist expression and may be valid in the regime where logs in  $Q^2$  dominate logs in  $x$ . The saturation scale for a fixed impact parameter is determined self-consistently by requiring that the dipole amplitude (within brackets in eq. 5) have the magnitude  $\mathcal{N}(x, r_S, \mathbf{b}_\perp) = 1 - e^{-1/2}$ , with  $Q_{s,p}^2 = 2/r_S^2$ . We note that there is an overall logarithmic uncertainty in the determination of  $Q_{s,p}^2(x, \mathbf{b}_\perp)$ .

For each value of the dipole radius, the gluon density  $xg(x, \mu^2)$  is evolved from  $\mu_0^2$  to  $\mu^2$  using LO DGLAP evolution equation without quarks,

$$\frac{\partial xg(x, \mu^2)}{\partial \log \mu^2} = \frac{\alpha_S(\mu^2)}{2\pi} \int_x^1 dz P_{gg}(z) \frac{x}{z} g\left(\frac{x}{z}, \mu^2\right) \quad (8)$$

Here the gluon splitting function with  $n_f$  flavor and  $C_A=3$  &  $T_R=1$  is

$$P_{gg}(z) = 6 \left[ \frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + \left( \frac{11}{2} - \frac{n_f}{3} \right) \delta(1-z) \quad (9)$$

The initial gluon density at the scale  $\mu_0^2$  is taken to be of the form

$$xg(x, \mu_0^2) = A_g x^{-\lambda_g} (1-x)^{5.6} \quad (10)$$

An important feature of the IP-Sat model is the b-dependence of the dipole cross-section, which is introduced through a gluon density profile function  $T(b)$ . This profile function is normalized to unity and is chosen to have the Gaussian form

$$T_p(\mathbf{b}_\perp) = \frac{1}{2\pi B_G} \exp \left( -\frac{\mathbf{b}_\perp^2}{2B_G} \right), \quad (11)$$

where  $B_G$  is a parameter fit to the HERA diffractive data. This corresponds to  $\langle b^2 \rangle = 2B_G$ , the average squared *gluonic* radius of the proton.

$m_c$	$B_G(\text{GeV}^{-2})$	$\mu_0(\text{GeV}^2)$	$A_g$	$\lambda_g$
1.4	4.0	1.17	2.55	0.020
1.35	4.0	1.20	2.51	0.024
1.5	4.0	0.77	2.64	0.011
1.4	4.0	1.50	3.61	-0.118

TABLE I: Parameters of the IP-Sat model obtained from the fit to HERA data [14].

Sets of parameters obtained from optimal fits of the IP-Sat model to HERA data [14] are listed in table I. All data sets except the last use  $m_{u,d,s} = 0.14$  GeV; the last set corresponds to  $m_{u,d,s} = 0.05$  GeV. The parameters of the initial gluon distribution are determined from fits to the HERA  $F_2$  data [27, 28] with a  $\chi^2 \sim 1$ . For charm quarks,  $x = x_{bj}(1 + 4m_c^2/Q^2)$ . The value of  $B_G$  is determined primarily from the  $t$ -distributions of  $J/\psi$  mesons measured by ZEUS [29] and H1 [30]. With these parameters, excellent agreement is obtained with the HERA exclusive vector meson and DVCS data. For a detailed comparison of this model to the HERA data, we refer the reader to Ref. [14].

### B. The b-CGC Model

The IP-sat dipole model is applicable when leading logarithms in  $Q^2$  dominate over leading logarithms in  $x$ . At very small  $x$ , quantum evolution in the CGC describing both the bremsstrahlung limit of linear small  $x$  evolution as well as nonlinear RG evolution at high parton densities, combined with a realistic  $b$ -dependence, may be better captured in the bCGC model [13–15]. The proton dipole cross-section in this case is expressed as

$$\frac{d\sigma_{\text{dip}}^p(\mathbf{r}_\perp, x, \mathbf{b}_\perp)}{d^2\mathbf{b}_\perp} = 2 \times \begin{cases} \mathcal{N}_0 \left( \frac{\mathbf{r}_\perp \bar{Q}_s}{2} \right)^{2\left(\gamma_s + \frac{1}{\kappa \lambda Y} \ln\left(\frac{2}{\mathbf{r}_\perp \bar{Q}_s}\right)\right)} & : \mathbf{r}_\perp \bar{Q}_s \leq 2; \\ 1 - \exp\left(-A \ln^2(B \mathbf{r}_\perp \bar{Q}_s)\right) & : \mathbf{r}_\perp \bar{Q}_s > 2; \end{cases} \quad (12)$$

In this model, in contrast to the IP-sat model, the impact parameter dependence is introduced through the quantity  $\bar{Q}_s(x, \mathbf{b}_\perp)$ , defined as

$$\bar{Q}_s(x, \mathbf{b}_\perp) = \left(\frac{x_0}{x}\right)^{\lambda/2} \left[ \exp\left(-\frac{\mathbf{b}_\perp^2}{2B_{\text{CGC}}}\right) \right]^{\frac{1}{2\gamma_s}} \quad (13)$$

As previously, for comparison of scales among different saturation models, the relevant saturation scale for a fixed impact parameter is determined self-consistently by requiring that the dipole amplitude (the expression to the right of the curly bracket in eq. 12) have the magnitude  $\mathcal{N}(x, r_S, \mathbf{b}_\perp) = 1 - e^{-1/2}$ , with the saturation scale defined as  $Q_{s,p}^2 = 2/r_S^2$ . The coefficients  $A$  and  $B$  are obtained by requiring the two asymptotic forms of the dipole cross-section and their first derivatives are continuous at  $\mathbf{r}_\perp \bar{Q}_s = 2$ :

$$A = \frac{\mathcal{N}_0^2 \gamma_s^2}{(1 - \mathcal{N}_0)^2 \ln(1 - \mathcal{N}_0)} \quad \& \quad B = \frac{1}{2} (1 - \mathcal{N}_0)^{-\frac{(1 - \mathcal{N}_0)}{\mathcal{N}_0 \gamma_s}} \quad (14)$$

The parameter  $\kappa = 9.9$  is fixed from the leading order BFKL value for this quantity

Table II presents the parameters of the model that are fitted to the HERA data [15]. The parameter  $B_{\text{CGC}}$  is determined from the  $t$ -dependence of exclusive  $J/\psi$  photo-production. For the b-CGC model, this parameter however cannot be easily interpreted as giving the square mean gluonic radius of the proton. The parameters presented in the table in the first and fourth lines do not give good fits to the data. The fit corresponding to the second line of the table gives the best fit to the data with  $\chi^2 \sim 1$ . The third line of the table corresponds to a fit where no saturation form is employed (namely, the perturbative expression, without the diffusion term proportional to  $Y = \ln(x_0/x)$ , is extended to  $\mathbf{r}_\perp \bar{Q}_s \geq 2$ ); it gives equally good fits to the data. However, it should be noted that this choice of parameters will violate perturbative unitarity for large dipole sizes  $\mathbf{r}_\perp > 1/\bar{Q}_s$ .

The dependence of the saturation scale  $Q_{s,p}^2(x, \mathbf{b}_\perp)$  as a function of  $x$  for different  $\mathbf{b}_\perp$  (and vice versa) in the IP-Sat and b-CGC models is shown in fig. 2. In both cases, the fits to the HERA data result in a semi-hard scale ( $Q_S^2 \gg \Lambda_{\text{QCD}}^2$ ) with decreasing  $x$  and  $b$  values probed in the collisions. The existence of such scales and their increase with energy is what validates the whole approach of treating high parton densities in weak coupling. It would of course be naive to interpret the extracted numerical value of  $Q_S$  as being precisely the scale that controls the running of the coupling. As is well known, the scale that controls the running of the coupling can differ considerably from this “bare” scale in a given scheme for any given process.

$\gamma_s$	$B_{\text{CGC}}(\text{GeV}^{-2})$	$\mathcal{N}_0$	$x_0$	$\lambda$
0.63	5.5	0.417	$5.95 \cdot 10^{-4}$	0.159
0.46	7.5	0.558	$1.84 \cdot 10^{-6}$	0.119
0.43	7.5	0.565	$1.34 \cdot 10^{-6}$	0.109
0.54	6.5	0.484	$3.42 \cdot 10^{-5}$	0.149

TABLE II: Parameters of the b-CGC model obtained from fits to HERA data [15]. The second row of parameters gives the best fit to HERA data.

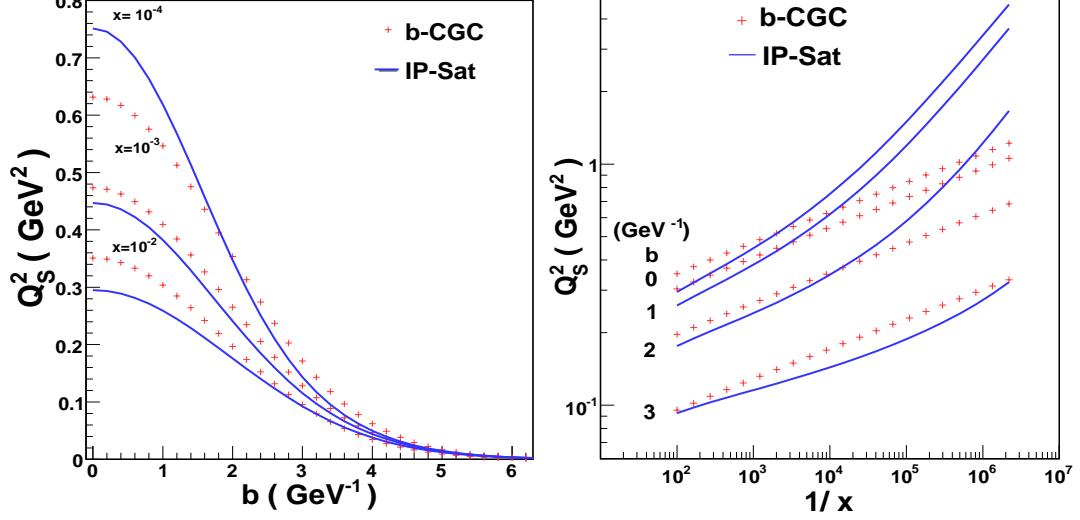


FIG. 2: Saturation scales obtained from the IP-Sat (blue lines) and b-CGC models (red crosses) using first set of parameters in table. I and the second set in table. II. Left: impact parameter dependence of saturation scale different values of  $x$ . Right:  $x$  dependence of saturation scale for different values of impact parameter  $b$ .

### C. The NLO-BK Model

The dipole cross-section is defined in eq. 2 as twice the forward scattering amplitude  $\mathcal{N}$ , which in the NLO-BK model [17] satisfies the equation,

$$\frac{\partial \mathcal{N}(\mathbf{r}_\perp, Y)}{\partial Y} = \int d\mathbf{r}_1 \mathcal{K}_{\text{NLO}}(\mathbf{r}_\perp, \mathbf{r}_1, \mathbf{r}_2) \times [\mathcal{N}(\mathbf{r}_1, Y) + \mathcal{N}(\mathbf{r}_2, Y) - \mathcal{N}(\mathbf{r}_\perp, Y) - \mathcal{N}(\mathbf{r}_1, Y)\mathcal{N}(\mathbf{r}_2, Y)], \quad (15)$$

where  $\mathbf{r}_2 \equiv \mathbf{r}_\perp - \mathbf{r}_1$  and the NLO-kernel (where running coupling corrections are taken into account) is given by

$$\mathcal{K}_{\text{NLO}}(\mathbf{r}_\perp, \mathbf{r}_1, \mathbf{r}_2) = \frac{\alpha_S(\mathbf{r}_\perp) N_c}{\pi} \left[ \frac{\mathbf{r}_\perp^2}{\mathbf{r}_1^2 \mathbf{r}_2^2} + \frac{1}{\mathbf{r}_1^2} \left( \frac{\alpha_S(\mathbf{r}_1^2)}{\alpha_S(\mathbf{r}_2^2)} - 1 \right) + \frac{1}{\mathbf{r}_2^2} \left( \frac{\alpha_S(\mathbf{r}_2^2)}{\alpha_S(\mathbf{r}_1^2)} - 1 \right) \right]. \quad (16)$$

This expression is based on considerable recent work to include running coupling corrections to the BK equation [31, 32]. It should be noted that the expression does not include other next-to-leading log contributions to the kernel that have been computed recently [33]. Also of relevance to us is the assumption in the evolution equation in eq. 15 that the dependence on the impact parameter and the dipole size factorize in the dipole amplitude as

$$\mathcal{N}(\mathbf{r}_\perp, x, \mathbf{b}_\perp) = 2 \mathcal{T}(\mathbf{b}_\perp) \mathcal{N}(\mathbf{r}_\perp, x). \quad (17)$$

The factorization here of impact parameter dependence and dipole size is very problematic conceptually and is an important limitation in applying these approaches to comparisons with data, except perhaps for final states that have limited sensitivity to the impact parameter dependence. How to include the impact parameter dependence in the BK/JIMWLK equations is an open question of great interest—we refer the reader to ref. [34] and references therein.

The NLO-BK model was applied in refs. [35] to a phenomenological study of the HERA data on the proton structure function  $F_2$ . Here the impact parameter dependence is taken to be a step function. Two different models for the

initial condition were used in that work. The first is the GBW model

$$N(r, Y = 0) = 1 - \exp \left[ -\frac{Q_{s_0}^2 r^2}{4} \right] \quad (18)$$

and the other is the MV model [2]

$$N(r, Y = 0) = 1 - \exp \left[ -\left( \frac{Q_{s_0}^2 r^2}{4} \right)^\gamma \ln \left( \frac{1}{r\Lambda_{QCD}} + e \right) \right] \quad (19)$$

The initial condition for protons was determined from a global fit of  $F_2$  data in the work of [35]. The fit parameters are summarized in the table III. The parameter  $C^2$  here is a scale that controls the running of the coupling in

I.C.	$\sigma_p$ (fm <sup>2</sup> )	$Q_{s_0,p}^2$ (GeV <sup>2</sup> )	$C^2$	$\gamma$
GBW	3.159	0.24	5.3	NA
MV	3.277	0.15	6.5	1.13

TABLE III: Parameters for the initial condition of the proton dipole cross section obtained in [35].

ref. [35]. This parametrization was extended to nuclei and fit to fixed target e+A DIS data [24]. It was then applied to study long range correlations in A+A collisions [24], single inclusive [21] and double inclusive [22] distributions in deuteron–gold collisions at RHIC. In this work, only the MV parametrization will be considered.

### III. SINGLE INCLUSIVE DISTRIBUTION IN P+P COLLISIONS

We shall now discuss the computation of the single inclusive hadron distribution in proton-proton collisions at the LHC obtained from the different dipole parametrizations discussed in the previous section. These parametrizations, when inserted in eqs. 4 and 3 respectively, give the single inclusive gluon distribution. While it might appear that all the parameters are fixed by the fits to the HERA data, a comparison with the single inclusive hadron distribution quickly makes clear that one requires further assumptions and parameters. In making comparisons to the collider inclusive hadron data, we shall discuss the sensitivity of our results to variations in these parameters. The additional assumptions and parameters are as follows:

- The dipole cross-sections are fit to HERA data for  $x \leq 0.01$ . One therefore needs to make an assumption for  $\phi(x, k_\perp, b_\perp)$  for larger  $x > x_0 = 0.01$  values that kinematic regions of the proton-proton data are sensitive to. We use the parametrization [11]

$$\phi(x, k_\perp, b_\perp) = \left( \frac{1-x}{1-x_0} \right)^\beta \left( \frac{x_0}{x} \right)^{\lambda_0} \phi(x_0, k_\perp, b_\perp), \quad x > x_0. \quad (20)$$

This parametrization of the large  $x$  unintegrated gluon distribution is motivated by quark counting rules [36] with fixed  $\beta = 4$  and the parameter  $\lambda_0$  which ranges from 0–0.2 in fits to the data.

- The single inclusive gluon distribution in eq. 3 has a logarithmic infrared divergence which can be regulated either by putting a cutoff on lower limit of  $p_\perp$  or replacing  $p_\perp$  by  $m_\perp = \sqrt{p_\perp^2 + m^2}$  where the mass is a free parameter. The single inclusive  $p_\perp$  distribution is sensitive to the choice of  $m$ , which is fixed to be the same for data at all energy ranges.
- Eq. 3 corresponds to the rapidity distribution of inclusive gluons at a fixed impact parameter. However, what is measured is the pseudo-rapidity distribution; the rapidity can be expressed in terms of the pseudo-rapidity most generally as

$$y(\eta, p_\perp, m) = \frac{1}{2} \left[ \frac{\sqrt{m^2 + p_\perp^2} \cosh^2 \eta + p_\perp \sinh \eta}{\sqrt{m^2 + p_\perp^2} \cosh^2 \eta - p_\perp \sinh \eta} \right]. \quad (21)$$

The single inclusive distribution with respect to the pseudo-rapidity therefore contains a Jacobian from the conversion of the expression with respect to the rapidity. We choose for economy of parameters the mass term in eq. 21 to be the same as the one that regulates the infrared divergence.

- The minimum-bias single inclusive gluon distribution is obtained from the expression

$$\frac{d\bar{N}_g}{d^2p_\perp dy} = \frac{\int d^2\mathbf{b}_\perp \frac{dN_g}{d^2p_\perp dy}(\mathbf{b}_\perp)}{\int d^2\mathbf{b}_\perp} \quad (22)$$

In the IP-Sat model, the proton profile does not change with energy. The transverse diffusion of the proton, often termed Gribov diffusion [37], is not fully accounted for by the diffusion of the unintegrated single inclusive gluon distribution because this growth does not automatically ensure the proper growth of the inelastic cross-section. Towards this end, we parametrize the maximum limit of  $b$ -integration to take the form  $b_{\max.} = b_0 + C \ln(s)$ . Fitting the available data on average  $dN/d\eta$  as a function of energy we can extract  $b_0$  and  $C$ . For first set of parameters in table I, the IP-Sat model gives  $b_0=5.17 \text{ GeV}^{-1}$  and  $C=0.19$  with a choice of the mass term  $=0.4 \text{ GeV}$  used in the Jacobian. This value of  $b_0$  is close to  $2b_{\text{rms.}}$  in the IP-Sat model, where  $b_{\text{rms.}}$  is the root mean square gluonic radius of the proton. The quantity  $\pi b_{\max}^2$ , which is the denominator of eq. 22 can be interpreted as being closely related to the inelastic cross section contributing to particle production. Fig. 3 shows the variation of  $\pi b_{\max}^2$  in the IP-Sat model as a function of collision energy. These numbers are in the ballpark of estimates of the inelastic cross-sections at the LHC [38]—they are however significantly higher than the values at lower energies. This is because the numbers for  $b_{\max.}$  are extracted from fits of eq. 22 to data; it is therefore also sensitive to the uncertainties in the numerator of eq. 22. A possible interpretation is that these uncertainties are larger at lower energies thereby leading to an overestimate of the inelastic cross-section.

An additional point with regard to fig. 3 is that one observes a  $\sim 15\%$  difference of  $\pi b_{\max}^2$  at the highest LHC energies and may thereby hope to constrain the parameter  $m$ . However, due to the uncertainty in distinguishing the genuine inelastic cross-section from the NSD cross-section, it is unlikely at present that the 15% difference can be definitive in that regard. A similar form of  $b_{\max}$  when used to fit average  $dN/d\eta$  for b-CGC model gives  $C \sim 0$ . This suggests that because the impact parameter dependence of the dipole cross-section in the b-CGC model is tied in with its  $x$  dependence (see eq. 12), the non-trivial relation of the two in this model may well approximate the physics of Gribov diffusion.

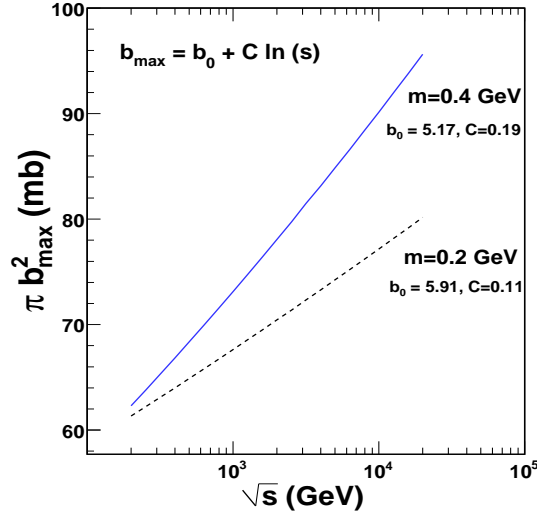


FIG. 3: Variation of the inelastic cross section with collision energy as obtained in the IP-Sat model for two different values of the infrared cut-off. See text for a detailed discussion.

- The single inclusive hadron distribution is obtained by convolving eq. 22 with the fragmentation function for gluons into charged hadrons<sup>1</sup>,

$$\frac{d\bar{N}_h}{d^2p_\perp dy} = \int_{z_{\min.}}^1 \frac{dz}{z^2} \frac{d\bar{N}_g}{d^2q_\perp dy} D_{g \rightarrow h} \left( z = \frac{p_\perp}{q_\perp}, \mu^2 \right), \quad (23)$$

<sup>1</sup> We thank A. Dumitru for pointing out an error in this expression in an earlier version of the paper. After correcting the error, the resulting expression gives better agreement with data.

where  $D_{g \rightarrow h}(z, \mu^2)$  is chosen to be  $6.05 z^{-0.714} (1 - z)^{2.92}$ , corresponding to the LO parameter set of [39]. The lower limit of the integral is determined from the kinematic requirement that  $x_{1,2} \leq 1$ . Our choice of fragmentation function is different from the fragmentation prescription of ref. [12] where the  $p_\perp$  of gluons is scaled by  $p_\perp / \langle z \rangle$ , with  $\langle z \rangle = 0.5$ . We also note that we do not have additional scales such as  $p_{\perp, \text{intrinsic}}$  in our comparison to data, in contrast to ref. [12].

### A. Results for the single inclusive hadron distribution

Now that we have specified all the assumptions and parameters that go into eq. 23, we are ready to present the results of our comparisons of the different saturation models with the data on rapidity and transverse momentum distributions for a wide range of collider energies up to the highest present LHC energy of  $\sqrt{s} = 7$  TeV and make projections for  $\sqrt{s} = 10, 14$  TeV.

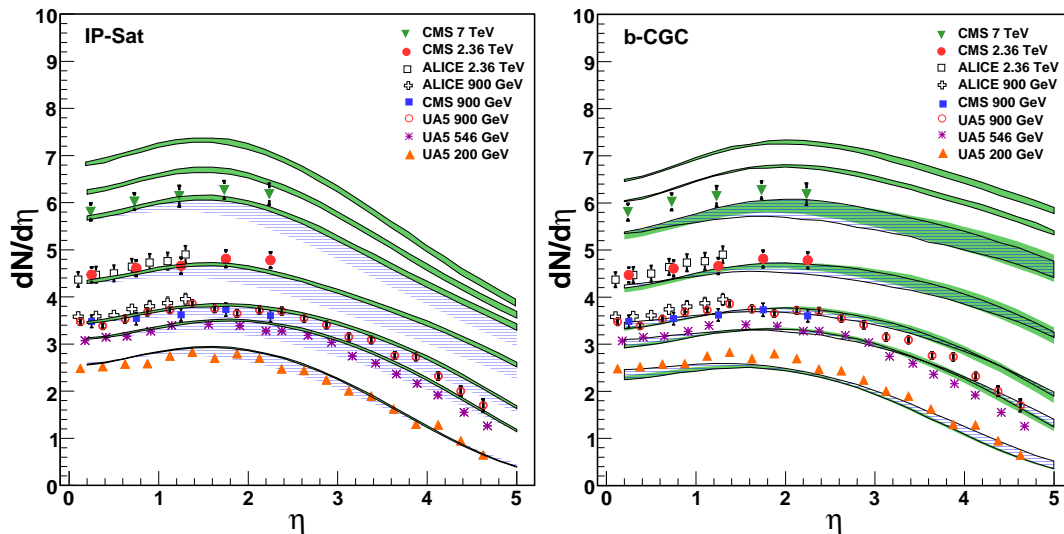


FIG. 4:  $dN/d\eta$  obtained from comparing IP-Sat and b-CGC models to data from UA5 [40], ALICE [54] and CMS [55]. The solid green band corresponds to uncertainties from different parameters; the dashed band is due to the variation of the choice of mass term in the Jacobian relating  $y$  to  $\eta$ . The two curves at the top in both panels correspond to projections in the two models for  $\sqrt{s} = 14$  and 10 TeV respectively.

Fig. 4 shows the pseudo-rapidity distribution obtained by integrating eq. 3 over  $p_\perp$  for different mass terms in the Jacobian and parameters given in the tables I and II for the IP-Sat and b-CGC dipole models respectively. The uncertainties corresponding to the choices of parametrizations and the infrared mass scale are shown in bands. In the IP-Sat model, the normalization is performed to the  $\sqrt{s}$  dependence of  $dN/d\eta$  at  $\eta = 0$  as shown in fig. 6—this fixes the two parameters in  $b_{\text{max}}$  we discussed previously. In the b-CGC model, because there is one less parameter (the coefficient of  $\ln s$  in  $b_{\text{max}}$  is zero), the normalization can be performed to the rapidity distribution at one fixed energy. In the figure shown, the normalization is performed for  $\sqrt{s} = 900$  GeV; choosing a lower energy corresponds to a  $< 10\%$  uncertainty in the overall normalization.

We didn't use a fragmentation function in computing the pseudo-rapidity distributions because the rapidity distribution is vastly dominated by contributions below  $p_\perp = 1$  GeV, where fragmentation functions are likely not reliable. We have varied the mass term in the Jacobian corresponding to eq. 21 in the range 0.2–0.4 GeV, corresponding to an infrared scale of order  $\Lambda_{QCD}$ . (In each case we chose this mass term to be equal to the one we use to regulate the infrared divergence of the unintegrated gluon distribution in eq. 4). The effect of the extrapolation parameter  $\lambda_0$  in eq. 20 is significant only at lower energies and higher values of  $\eta$ . The agreement with data of both models is quite good with the IP-Sat model providing a somewhat better agreement at the highest energies.

In fig. 5, the corresponding  $p_\perp$  distributions, with the previously specified fragmentation prescription, is shown. The IP-Sat model shows a poor agreement with data for the lower energies at high  $p_\perp$ , as does the b-CGC model with some of the HERA parameter sets. At higher  $p_\perp$ , the results are sensitive to physics at  $x \geq 0.01$ , which is parametrized very simply in the models. On the reverse side of the coin, one should anticipate a better agreement in the same  $p_\perp$  window at higher energies. Indeed, a systematically better agreement is seen in both models with the data at the higher energies for all parameter sets.



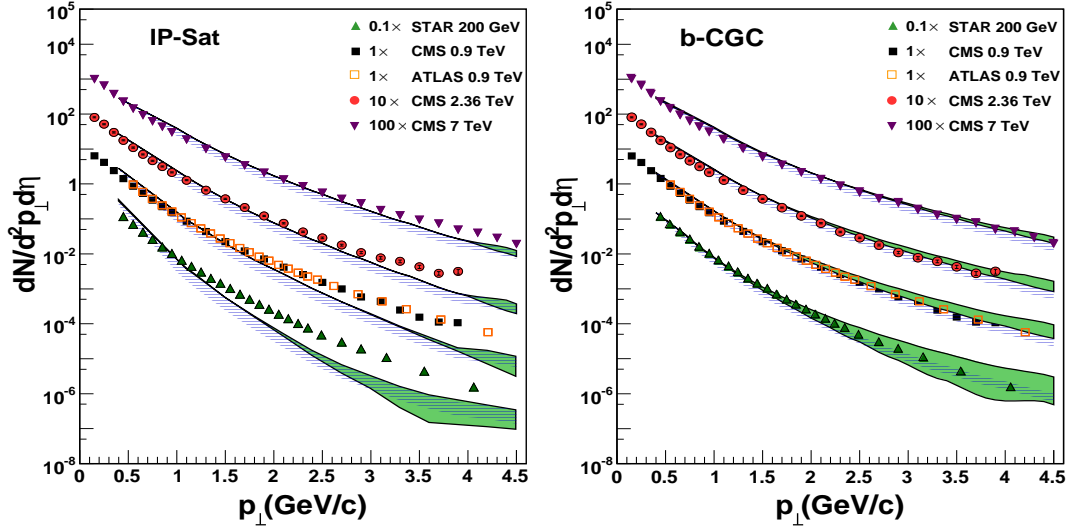


FIG. 5:  $dN/d^2p_\perp d\eta$  in the IP-Sat and b-CGC models. The solid green band corresponds to uncertainties from different parameters and the dashed band is due variation of the mass term in the Jacobian relating  $\eta$  to  $y$ . The  $p_\perp$  distribution is averaged over the  $\eta$  range of  $\pm 2.4$ . The experimental data points are from CMS [55], STAR [56] and ATLAS [57]

Fig. 6 shows the average value of  $dN/d\eta$  calculated at  $\eta=0$ . The left plot of fig. 6 shows a fit of  $dN/d\eta$  using different functional forms for the saturation scales. We considered both  $Q_S^2$  and  $Q_S^2/\alpha_S(Q_S)$ . In the CGC framework, one would expect the latter. However, if the running is not significant in the energy range of interest, the former is also possible. Indeed, a good fit to the CMS data was obtained [44] with a simple form of the saturation scale *a la* the Golec-Biernat model [7]. In our case, the comparison is made within the framework of the IP-Sat and b-CGC models which give better fits to the HERA data and are sensitive to the impact parameter profile of the gluon distribution in the proton. For the comparison with the LHC data in the left panel of fig. 6, we use the value of  $Q_S$  at the median value of  $s_\perp=2$  GeV $^{-1}$ . The dependence of  $dN/d\eta$  on the purely  $Q_S^2$  functional form is not very good, while the  $Q_S^2/\alpha_S(Q_S)$  form does much better for the IP-Sat model. For the running of  $\alpha_S$ , we chose  $Q_S^2(s_\perp)$  at  $s_\perp=0$  to restrict its running to  $\alpha_S$  below 0.5. Fig. 6 (right panel) shows by way of comparison, a comparison of  $dN/d\eta$  at  $\eta=0$  as a function of  $\sqrt{s}$  to IP-Sat and b-CGC models. In the IP-Sat model, a good fit is ensured because a fit to this energy dependence is what determines the parameters of  $b_{\text{max}}$ —see fig. 3 and related discussion. In the b-CGC model, the curve is a prediction and is seen to be a very good fit to the data.

The energy dependence of  $\langle p_\perp \rangle$  is shown in fig. 7. In the left panel, it is shown that in this case one obtains a good linear dependence of  $\langle p_\perp \rangle$  on  $Q_S$  in both the IP-Sat and b-CGC models as the c.m energy is varied. This is as seen previously [44]. The right plot shows  $\langle p_\perp \rangle$  versus  $\sqrt{s}$  computed in the IP-Sat and b-CGC models. Here one sees that the results are quite sensitive to choice of the infrared cut-off.

Fig. 8 shows the  $\eta$  and  $p_\perp$  distributions computed in the NLO-BK model. Only MV initial conditions are considered. In this model, the impact parameter dependence of the inclusive gluon multiplicity is given by

$$\mathcal{T}(b) = \frac{1}{\pi b_{\text{max}}^2} \Theta(b_{\text{max.}} - b) .$$

Therefore  $d\bar{N}_g(b)/d\eta = d\bar{N}_g/d\eta \mathcal{T}(b)$ , where  $b_{\text{max.}}$  is a parameter that can be absorbed in the normalization. If no dependence of  $b_{\text{max.}}$  on  $\sqrt{s}$  is assumed, the model considerably overestimates the single inclusive data at LHC energies. This is likely a consequence of the fact that the NLO-BK doesn't take into account the impact parameter dependence of the saturation scale. In this case, the values for the inclusive gluon multiplicity correspond to the values for the zero impact parameter, which is considerably higher than the minimum bias values. The agreement with data is improved considerably by allowing  $b_{\text{max.}}$  to depend on  $\sqrt{s}$ . The denominator  $\pi b_{\text{max.}}^2$  therefore provides an energy dependent normalization. The same approach discussed previously for the IP-Sat model is employed to extract the  $\sqrt{s}$  dependence of  $b_{\text{max.}}$  chosen to be of the form  $b_{\text{max.}} = b_0 + C \ln(s)$ . From a fit to the average  $dN/d\eta$  as a function of energy, one obtains  $b_0 \sim 5.6 - 7.55$  GeV $^{-1}$  and  $C \sim 0.23 - 0.46$  depending on the choice of infrared cut-off.

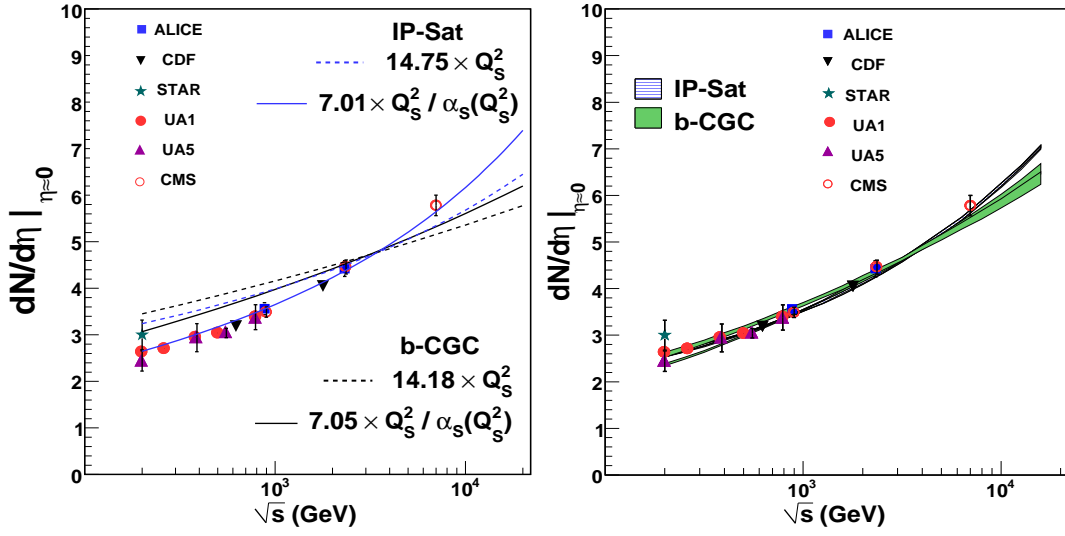


FIG. 6: Average  $dN/d\eta|_{\eta=0}$  in the IP-Sat and b-CGC models. Left: Data plotted as a function of saturation scales (at median impact parameter  $b = 2 \text{ GeV}^{-1}$ ) for both the models determined from the HERA data. Right: Average  $dN/d\eta$  at  $\eta = 0$  from the  $k_\perp$ -factorized expression in eq. 23 from IP-Sat (solid green) and b-CGC (dashed) models. See text for further discussion. Experimental data points are from Ref.[55], [60], [61], [62], [56], [58].

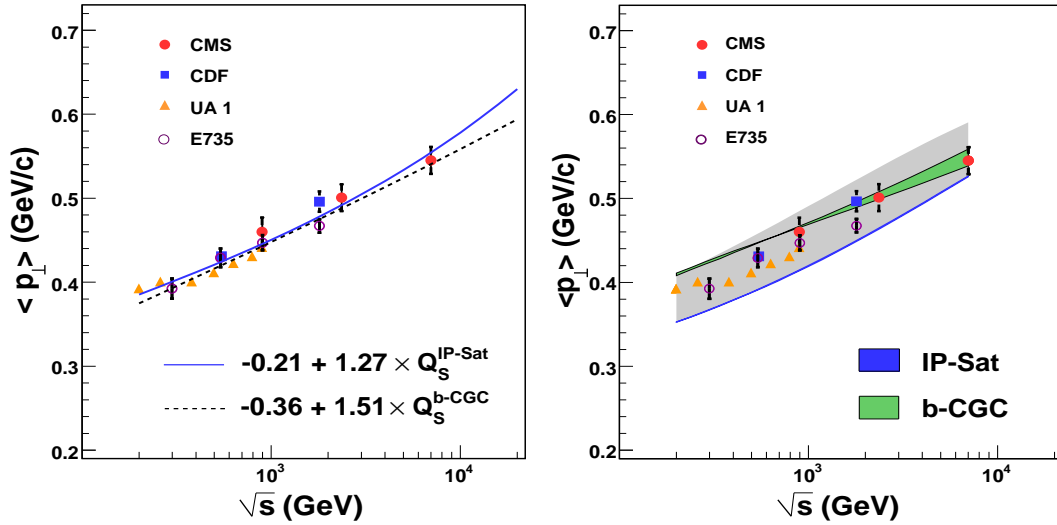


FIG. 7: Average  $p_\perp$  obtained from IP-Sat and b-CGC models compared to data. Left: Function of saturation scales fitted for both the models. Right: Average  $p_\perp$  from the  $k_\perp$ -factorized expression in eq. 23. The thin (colored) bands correspond to uncertainties arising from different parameters in table I and table II and mass term=0.2 GeV. The thick (gray) band shows the sensitivity to variation of mass term in the range 0.2–0.3 GeV (with lower mass corresponding to lower  $\langle p_\perp \rangle$ ). Experimental data points are from Ref. [55] [63] [62] [61].

#### IV. MULTIPLICITY DISTRIBUTION

There are several sources of multiplicity fluctuations in high energy hadronic collisions. These can arise from fluctuations in the number of wee partons, in their distribution with impact parameter and their distribution in rapidity [45]. In this paper, we will consider particle production in a relatively small rapidity window (parametrically of order  $\Delta\eta \leq 1/\alpha_s$ ), so fluctuations in rapidity will not be an important source of fluctuations. Let us first consider fluctuations in multiplicity for a fixed impact parameter. In this case, the CGC framework allows for a systematic treatment of inclusive multi-particle production [47] in the Glasma [46]. The largest contribution to multi-particle production comes from diagrams that appear superficially disconnected, but are connected by averaging over color correlations in an event and over all events. This is the formal basis of the Glasma flux tube picture [48], and was

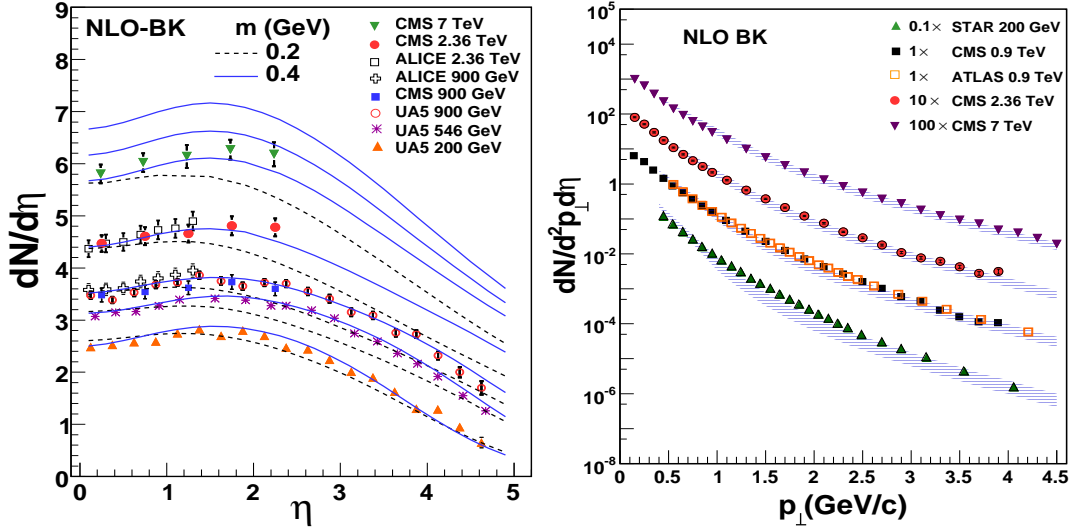


FIG. 8: Pseudo-rapidity and  $p_\perp$  distribution in the NLO-BK model compared to data. The uppermost two plots in the left panel correspond to predictions for  $\sqrt{s} = 14, 10$  TeV with  $m = 0.4$  GeV. The  $p_\perp$  distribution is averaged over the  $\eta$  range of  $\pm 2.4$ . The band corresponds to the variation  $m = 0.2$ – $0.4$  GeV in the Jacobian relating  $\eta$  to  $y$ .

previously used to compute backward-forward correlations [49, 50], two particle correlations [24, 48], three particle correlations [51], and n-particle correlations [52]. It is this last computation that will concern us here. The n-particle correlations obtained by averaging over color sources in the CGC picture are those that would be generated by the negative binomial distribution [52]. To the best of our knowledge, this derivation of the negative binomial distribution as arising from particle production from Glasma flux tubes is the first such *ab initio* derivation in a QCD framework<sup>2</sup>

The negative binomial distribution is given by

$$P_n^{\text{NB}}(\bar{n}, k) = \frac{\Gamma(k+n)}{\Gamma(k)\Gamma(n+1)} \frac{\bar{n}^n k^k}{(\bar{n}+k)^{n+k}}. \quad (24)$$

The distribution is characterized by two parameters, the mean multiplicity  $\bar{n}$  and the parameter  $k$ . As is well known, in the limit  $k \rightarrow \infty$ , this distribution reduces to the Poisson distribution. In the limit  $k \rightarrow 1$ , one obtains the Bose-Einstein distribution. The variance of the distribution is given by  $\sigma^2 = \bar{n}^2 - \bar{n}^2 = \bar{n} + \bar{n}^2/k$ . In the Glasma flux tube approach,  $k$  is not an arbitrary parameter; instead, it is computed to be

$$k = \zeta \frac{(N_c^2 - 1) Q_S^2 S_\perp}{2\pi}, \quad (25)$$

where  $\zeta$  is a dimensionless non-perturbative parameter which will be discussed shortly and  $S_\perp$  is the overlap area of the two hadrons.

As we mentioned previously, the negative binomial distribution is obtained at a fixed impact parameter, so  $\bar{n} \equiv \bar{n}(b)$  and  $k \equiv k(b)$ . In particular, the latter parameter must be interpreted as being proportional to the number of flux tubes (or interacting “hot spots”)  $S_\perp/1/Q_S^2$  at a given impact parameter. Because  $Q_S^2$  grows with energy,  $k$  has a very particular energy dependence. Here, the parameter  $k$  in eq. 25 is determined as follows. For a given impact parameter, we define

$$Q_S^2 S_\perp = \int d^2x_\perp Q_S^2(x_\perp),$$

where the integral on the r.h.s is performed over the overlap area of the two protons at a given parameter. At each given transverse position  $x_\perp$  in the overlap area, we choose  $Q_S(x_\perp) = \min.\{Q_S^A, Q_S^B\}$ , where  $Q_S^A$  and  $Q_S^B$  are

<sup>2</sup> The negative binomial distribution has of course been known for a long time [64] to provide good fits to collider p+p data. More recently shown to describe the multiplicity distributions in A+A collisions at RHIC [65]. However, a microscopic derivation of this distribution from QCD was previously lacking.

respectively the saturation scales of the two colliding protons at that  $x_\perp$ . This choice is motivated by the fact that the inclusive multiplicity of produced gluons is much more sensitive to the smaller of the two saturation scales [53]. In fig. 9 (left) we plot the saturation scale as a function of  $\sqrt{s}$  for different impact parameters. The right plot has on

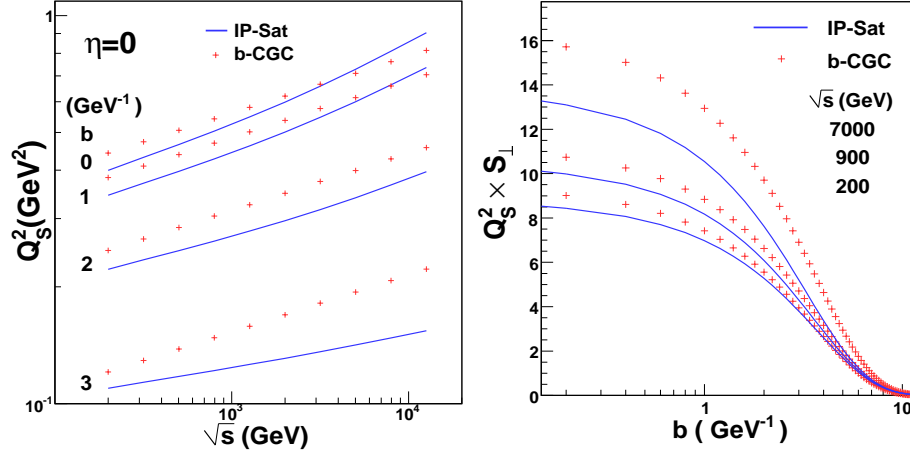


FIG. 9: Left: Variation of the saturation scale with the collision energy. Right: Variation of  $Q_s^2 S_\perp$  for different impact parameters and the c.m. energy of the collision. The solid lines are for the IP-Sat model and crosses for the b-CGC model, in each case for the parameters providing the best fit to the HERA data.

the  $y$ -axis the quantity that appears in eq. 25 determined by the procedure we described. We observe that a stronger dependence of this quantity is seen for the b-CGC model. With the exception of the parameter  $\zeta$ , we have everything necessary to compute eq. 24 at a given impact parameter.

Fluctuations in impact parameter are treated as follows. The overlap function for two protons at a given impact parameter (see fig. 1 (right)) can be expressed as

$$T_{pp}(\mathbf{b}_\perp) = \int d^2\mathbf{s}_\perp T_p(\mathbf{s}_\perp) T_p(\mathbf{s}_\perp - \mathbf{b}_\perp) \quad (26)$$

where  $T_p$  is given for instance in the IP-Sat model by eq. 11. Our knowledge of the HERA diffractive data therefore allows one to compute  $T_{pp}$  in the saturation models. The probability distribution for an inelastic collision at a given impact parameter is given in impact parameter eikonal models as [38, 41, 42]

$$\frac{dP_{\text{inel.}}^{\text{eik.}}}{d^2\mathbf{b}_\perp} = \frac{1 - \exp(-\sigma_{gg} T_{pp})}{\int d^2\mathbf{b}_\perp (1 - \exp(-\sigma_{gg} T_{pp}))}. \quad (27)$$

In general,  $\sigma_{gg}$  is an energy dependent quantity estimated to be the elementary cross-section for gluon-gluon scattering. Alternately, in our framework, an estimate for this quantity in our framework involving no additional parameters is

$$\frac{dP_{\text{inel.}}^{\text{dip.}}}{d^2\mathbf{b}_\perp} = \frac{\frac{dN_g}{dy}(\mathbf{b}_\perp)}{\int d^2\mathbf{b}_\perp \frac{dN_g}{dy}(\mathbf{b}_\perp)}. \quad (28)$$

This expression of course gives unity when both sides are integrated over impact parameter. The result for  $2\pi b \, dP_{\text{inel.}}^{\text{dip.}}/d^2\mathbf{b}_\perp$  at a fixed  $\sqrt{s} = 900$   $\text{GeV}$  is shown in fig.10 and is a sharply peaked distribution at  $\sim 3$   $\text{GeV}^{-1}$ . The distribution plotted is insensitive to  $\sqrt{s}$  and to the infrared scale  $m$ . We should note here that the inclusive multiplicity at a given impact parameter is given by  $\bar{n}(\mathbf{b}_\perp) = P_1 + 2P_2 + 3P_3 + \dots$ , while the inelastic probability is given by  $P_{\text{inel.}} = P_1 + P_2 + P_3 + \dots$ . In general, the two are of course not the same, so using eq. 28 to estimate the inelastic probability is strictly not correct. Our justification here is that eq. 28 has the right qualitative behavior and involves one less parameter than eq. 27. It is very desirable to obtain a better estimate for eq. 28. For a detailed discussion of the problem of computing multi-particle production in the presence of strong time dependent sources, see ref. [43].

With the stated assumptions (and caveats) we are now in a position to compute the probability distribution as a function of multiplicity. By convolving the probability distribution for producing  $n$  particles at a given impact parameter (eq. 24) with the probability for an inelastic collision at that impact parameter (from eq. 28), one obtains

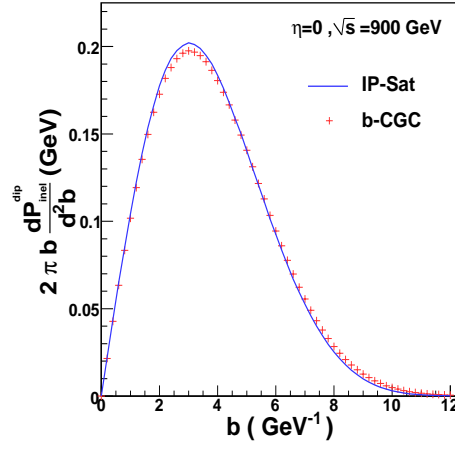


FIG. 10: Probability distribution as a function of impact parameter for an inelastic collision computed using eq. 28 for the b-CGC and IP-Sat models.

the expression

$$P(n) = \int d^2\mathbf{b}_\perp \frac{dP_{\text{inel.}}}{d^2\mathbf{b}_\perp} P_n^{\text{NB}}(\bar{n}(\mathbf{b}_\perp), k(\mathbf{b}_\perp)) \quad (29)$$

The results for this quantity are shown in fig. 11. Note that since the input here is the average inclusive multiplicity at a given impact parameter (as opposed to the minimum bias inclusive multiplicity that was compared to data), this quantity needs to be normalized as well. We do so by fitting the multiplicity distribution corresponding to the lowest energy UA5 data set at  $\sqrt{s} = 200$  GeV for the normalization of  $\bar{n}$ . In the b-CGC model, a fit of the overall normalization to the single inclusive minimum bias distribution at a given energy gives the same value (to an accuracy of  $< 5\%$ ) as that obtained for the same quantity if fit to the multiplicity distribution instead. In the IP-Sat model, one obtains the same normalization constant if  $b_{\text{max.}} = 2b_{\text{rms.}}$ . If one chooses the form  $b_{\text{max.}} = b_0 + C \ln(s)$  we described previously, there is a 60% discrepancy between the two choices of fixing the normalization. The agreement between data and model is remarkably good for IP-Sat distribution for all energies and rapidity cuts<sup>3</sup>. For the b-CGC model, the agreement with data for  $|\eta| < 0.5$  is quite good for the 2.36 TeV ALICE data [60] but shows deviations for other energies at the highest multiplicities<sup>4</sup>.

An important point regarding the comparison of the models to data in fig. 11 concerns the magnitude of the parameter  $\zeta$  in eq. 25 which is fit to the data. In the b-CGC model, it is extracted to be 0.25 and it is 0.35 in the IP-Sat model<sup>5</sup>. This quantity was recently computed non-perturbatively *ab initio* by solving the Yang-Mills equations numerically [66] for two gluon correlations from gauge fields generated in the collision of two dense color sources [66]. The results of the numerical computation vary depending on parameter choices in the range  $\zeta \sim 0.3$ -1.5—the essential point is that the result is a number of order unity. It is very encouraging that the values extracted from the data for  $\zeta$  in the saturation models are numbers of this order<sup>6</sup>, and suggests that careful analysis of the data can help to extract non-trivial information about the screening of Glasma flux tubes in p+p collisions.

<sup>3</sup> The results are insensitive to the choice of the infrared cut-off  $m$ .

<sup>4</sup> Our computation of the saturation scale is performed for  $\eta = 0$  while the data we compare with are averaged over  $|\eta| < 0.5$  and  $|\eta| < 1$ . This discrepancy may lead to small corrections to our results.

<sup>5</sup> We have not attempted any fine tuning of our fits at this stage. A good agreement can be obtained for the 2.36 TeV and 7 TeV data in the b-CGC model by decreasing  $\zeta$  by 20% but this gives a significantly worse fit at lower energies. However, because the model's validity is questionable at lower energies, and because of other sources of uncertainty we have articulated, the b-CGC model still provides a “competitive” mechanism for multi-particle production.

<sup>6</sup> Nothing in principle would have prevented the number for  $\zeta$  extracted from the data from being orders of magnitude off, in which case the present analysis would be invalidated completely.

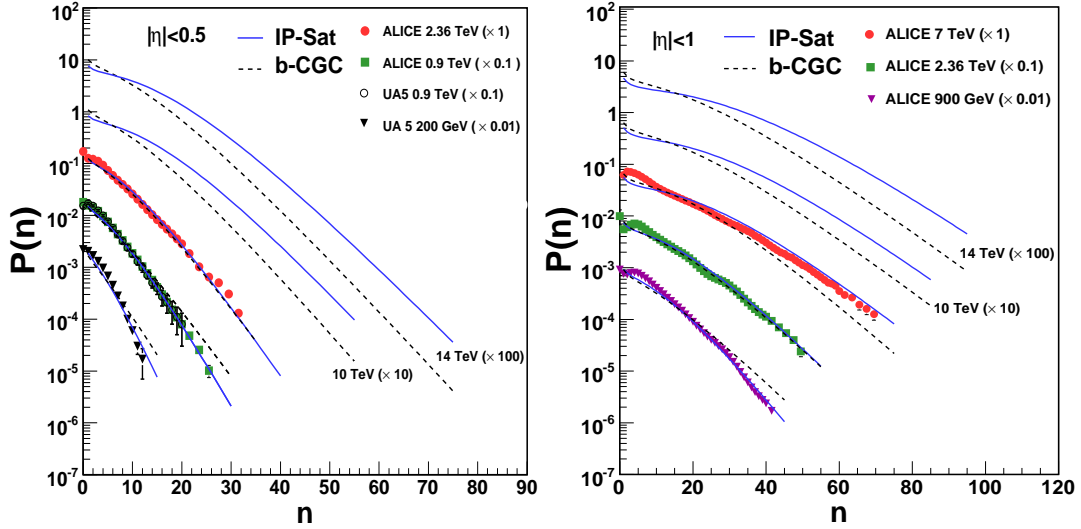


FIG. 11: Probability distribution of the gluon multiplicity computed in the IP-Sat model (parameter set I) and the b-CGC model (parameter set II) compared to the UA5[59] and ALICE[60] data for different  $\eta$  ranges. Left: Multiplicity distribution for the  $\eta$  range  $\pm 0.5$ . Right: Multiplicity distribution for the  $\eta$  range  $\pm 1$ .

## V. SUMMARY AND OUTLOOK

In this work, we extracted impact parameter dependent unintegrated gluon distributions from fits to the HERA inclusive and exclusive data in the IP-Sat, b-CGC and NLO-BK models. These models all implement the physics of saturation but differ in their dynamical assumptions. The impact parameter distributions, in the  $k_\perp$  factorization formalism allow one to compute single inclusive rapidity and  $p_\perp$  distributions in p+p collisions at the LHC. These give quite reasonable agreement with the LHC data. These impact parameter dependent distributions also allowed us to compute the multiplicity distribution, which in the Color Glass Condensate/Glasma formalism is predicted to be a negative binomial distribution with particular values for the parameter controlling the width of the distribution. We observed that the multiplicity distributions are well reproduced in this framework with numbers for the parameter consistent with values extracted from numerical computations of Yang-Mills equations for the collision of dense color sources. These results suggest that particle emission from Glasma flux tubes generated in collisions of “hot spots” of size  $1/Q_S$  are a strong candidate for generating the multi-parton correlations underlying the multiplicity distribution. Because these Glasma flux tubes generate long range rapidity correlations, our results are also consistent with a recent claim [68] that Glasma flux tubes generate the near side ridge seen in high multiplicity events by the CMS collaboration [67]. Because our analysis allows a more careful treatment of the contribution of various impact parameters to the multiplicity distribution, it will allow more quantitative and systematic comparisons to phenomena seen only in high multiplicity cuts in the LHC data.

There are several caveats in this analysis that must be noted. Firstly, the  $k_\perp$  factorization formalism for inclusive multiplicity distributions is rather fragile for  $k_\perp \leq Q_S$  because it does not include multi-parton rescatterings [69, 70]. Significant improvements of the  $k_\perp$  factorization formalism are feasible in the CGC framework if presently cumbersome to implement in phenomenological analyses. Similarly, NLO dipole computations are becoming available albeit a consistent NLO dipole analysis still remains to be developed. Both these developments suggest that a combined quantitative study of the collective QCD dynamics of saturation in DIS and hadronic collisions is feasible in future along the lines discussed in this work.

## Acknowledgements

R.V was supported by the US Department of Energy under DOE Contract No.DE-AC02-98CH10886. We thank Kevin Dusling and François Gelis for a careful reading of the manuscript and especially Adrian Dumitru for pointing out an error in a previous version of the manuscript. We gratefully acknowledge useful conversations with Javier

- 
- [1] L.V. Gribov, E.M. Levin, M.G. Ryskin, Phys. Rept. **100**, 1 (1983); A.H. Mueller, J-W. Qiu, Nucl. Phys. **B 268**, 427 (1986).
  - [2] L.D. McLerran, R. Venugopalan, Phys. Rev. **D 49**, 2233 (1994); *ibid.* **49**, 3352 (1994); *ibid.* **50**, 2225 (1994).
  - [3] Jalilian-Marian, A. Kovner, A. Leonidov, H. Weigert, Nucl. Phys. **B 504**, 415 (1997); *ibid.*, Phys. Rev. **D 59**, 014014 (1999); E. Iancu, A. Leonidov, L.D. McLerran, Nucl. Phys. **A 692**, 583 (2001); E. Ferreira, E. Iancu, A. Leonidov, L.D. McLerran, Nucl. Phys. **A 703**, 489 (2002).
  - [4] E. Iancu, R. Venugopalan, Quark Gluon Plasma 3, Eds. R.C. Hwa, X.N. Wang, World Scientific, hep-ph/0303204; H. Weigert, Prog. Part. Nucl. Phys. **55**, 461 (2005); F. Gelis, E. Iancu, J. Jalilian-Marian, R. Venugopalan, arXiv:1002.0333.
  - [5] L. D. McLerran, R. Venugopalan, Phys. Rev. **D 59**, 094002 (1999); R. Venugopalan, Acta Phys. Polon. **B 30**, 3731 (1999).
  - [6] N. N. Nikolaev, B. G. Zakharov, Z. Phys. **C 49**, 607 (1991); A. H. Mueller, Nucl. Phys. **B 335**, 115 (1990).
  - [7] K. J. Golec-Biernat, M. Wusthoff, Phys. Rev. **D 59**, 014017 (1998); K. J. Golec-Biernat, M. Wusthoff, Phys. Rev. **D 60**, 114023 (1999).
  - [8] I. Balitsky, Nucl. Phys. **B 463**, 99 (1996); Yu.V. Kovchegov, Phys. Rev. **D 61**, 074018 (2000).
  - [9] J.P. Blaizot, F. Gelis, R. Venugopalan, Nucl. Phys. **A 743**, 57 (2004).
  - [10] M.A. Braun, Phys. Lett. **B 483**, 105 (2000).
  - [11] F. Gelis, A. M. Stasto, R. Venugopalan, Eur. Phys. J. **C 48**, 489 (2006).
  - [12] E. Levin, A. H. Rezaeian, Phys. Rev. **D 82**, 014022 (2010).
  - [13] E. Iancu, K. Itakura, S. Munier, Phys. Lett. **B 590**, 199 (2004).
  - [14] H. Kowalski, L. Motyka, G. Watt, Phys. Rev. **D 74**, 074016 (2006).
  - [15] G. Watt, H. Kowalski, Phys. Rev. **D 78**, 014016 (2008).
  - [16] H. Kowalski, D. Teaney, Phys. Rev. **D 68**, 114005 (2003).
  - [17] J. L. Albacete, Y. V. Kovchegov, Phys. Rev. **D 75**, 125021 (2007).
  - [18] J. R. Forshaw, R. Sandapen, G. Shaw, JHEP **0611**, 025 (2006); B. Z. Kopeliovich, J. Raufeisen, A. V. Tarasov, Phys. Rev. **C 62**, 035204 (2000); . Gotsman, E. Levin, M. Lublinsky, U. Maor, Eur. Phys. J. **C 27**, 411 (2003); G. Soyez, Phys. Lett. **B 655**, 32 (2007).
  - [19] D. Kharzeev, Yu. Kovchegov, K. Tuchin, Phys. Rev. **D 68**, 094013 (2003); A. Dumitru, A. Hayashigaki, J. Jalilian-Marian, Nucl. Phys. **A 770**, 57 (2006).
  - [20] V. P. Goncalves, M. S. Kugeratski, M. V. T. Machado, F. S. Navarra, Phys. Lett. **B 643**, 273 (2006).
  - [21] J. L. Albacete, C. Marquet, Phys. Lett. **B 687**, 174 (2010).
  - [22] J. L. Albacete, C. Marquet, Phys. Rev. Lett. **105**, 162301 (2010).
  - [23] H. Kowalski, T. Lappi, R. Venugopalan, Phys. Rev. Lett. **100**, 022303 (2008).
  - [24] K. Dusling, F. Gelis, T. Lappi, R. Venugopalan, Nucl. Phys. **A 836**, 159 (2010).
  - [25] T. Rogers, V. Guzey, M. Strikman, X. Zu, Phys. Rev. **D 69**, 074011 (2004).
  - [26] J. Bartels, K. J. Golec-Biernat, H. Kowalski, Phys. Rev. **D 66**, 014001 (2002).
  - [27] S. Chekanov *et al.* [ZEUS Collaboration], Eur. Phys. J. **C 21**, 443 (2001).
  - [28] C. Adloff *et al.* [H1 Collaboration], Eur. Phys. J. **C 21**, 33 (2001).
  - [29] S. Chekanov *et al.* (ZEUS Collaboration), Eur. Phys. J. **C 24**, 345 (2002), Nucl. Phys. **B695**, 3 (2004).
  - [30] A. Aktas *et al.* (H1 Collaboration), Eur. Phys. J. **C 46**, 585 (2006).
  - [31] I. Balitsky, Phys. Rev. **D 75**, 014001 (2007).
  - [32] Y. V. Kovchegov, H. Weigert, Nucl. Phys. **A 784**, 188 (2007).
  - [33] I. Balitsky, G. A. Chirilli, Phys. Rev. **D 77**, 014019 (2008); G. Beuf, arXiv:1008.0498 [hep-ph].
  - [34] J. Berger, A. Stasto, arXiv:1010.0671 [hep-ph].
  - [35] J. L. Albacete, N. Armesto, J. G. Milhano, C. A. Salgado, Phys. Rev. **D 80**, 034031 (2009).
  - [36] V. A. Matveev, R. M. Muradian, A. N. Tavkhelidze, Lett. Nuovo Cim. **7**, 719 (1973); S. J. Brodsky, G. R. Farrar, Phys. Rev. Lett. **31**, 1153 (1973).
  - [37] V. N. Gribov, arXiv:hep-ph/0006158.
  - [38] D. d’Enterria, G. K. Eyyubova, V. L. Korotkikh, I. P. Lokhtin, S. V. Petrushanko, L. I. Sarycheva, A. M. Snigirev, Eur. Phys. J. **C 66**, 173 (2010).
  - [39] B. A. Kniehl, G. Kramer, B. Potter, Nucl. Phys. **B 582**, 514 (2000).
  - [40] S. Eidelman *et al.* [Particle Data Group Collaboration], “Review of particle physics”, Phys. Lett. **B592** 1 (2004).
  - [41] L. Frankfurt, M. Strikman, C. Weiss, Ann. Rev. Nucl. Part. Sci. **55**, 403 (2005).
  - [42] M. L. Miller, K. Reygers, S. J. Sanders, P. Steinberg, Ann. Rev. Nucl. Part. Sci. **57**, 205 (2007).
  - [43] F. Gelis, R. Venugopalan, Nucl. Phys. **A 776**, 135 (2006); *ibid.* **A 779**, 177 (2006).
  - [44] L. McLerran, M. Praszalowicz, Acta Phys. Polon. **B 41**, 1917 (2010).
  - [45] H. I. Miettinen, J. Pumplin, Phys. Rev. **D 18**, 1696 (1978).
  - [46] T. Lappi, L.D. McLerran, Nucl. Phys. **A 772**, 200 (2006).
  - [47] F. Gelis, T. Lappi, R. Venugopalan, Phys. Rev. **D 78**, 054019 (2008); *ibid.*, **78**, 054020 (2008); *ibid.*, **79**, 094017 (2009).
  - [48] A. Dumitru, F. Gelis, L. McLerran, R. Venugopalan, Nucl. Phys. **A 810**, 91 (2008).

- [49] N. Armesto, L. McLerran, C. Pajares, Nucl. Phys. A **781**, 201 (2007).
- [50] T. Lappi, L. McLerran, Nucl. Phys. A **832**, 330 (2010).
- [51] K. Dusling, D. Fernandez-Fraile and R. Venugopalan, Nucl. Phys. A **828**, 161 (2009) [arXiv:0902.4435 [nucl-th]]
- [52] F. Gelis, T. Lappi, L. McLerran, Nucl. Phys. A **828** (2009) 149.
- [53] A. Dumitru, L. D. McLerran, Nucl. Phys. A **700**, 492 (2002).
- [54] ALICE Collaboration, Eur. Phys. J. C **65** (2010) 111 [arXiv:0911.5430]; arXiv:1004.3034.
- [55] CMS Collaboration, Phys Rev Lett **105**, 022002 (2010).
- [56] STAR Collaboration, Phys. Rev. Lett. **91**, 172302 (2003), Phys. Rev. C **79**, 034909 (2009).
- [57] ATLAS Collaboration, Phys. Lett. B **688** (2010) 21 [arXiv:1003.3124].
- [58] UA5 collaboration, Z. Phys. C **33** (1986) 1.
- [59] UA5 Collaboration, Z. Phys. C **43** (1989) 357.
- [60] ALICE Collaboration, Eur.Phys.J.C **68**:345-354,2010 [arXiv:1004.3514].
- [61] UA1 collaboration, A. M. Rossi et al., Nucl. Phys. B **84** (1975) 269.
- [62] CDF collaboration, Phys. Rev. D **41** (1990) 2330; Phys. Rev. Lett. **61** (1988) 1819.
- [63] E735 Collaboration, T. Alexopoulos et al., Phys. Rev. Lett. **60**, 1622 (1988).
- [64] E. A. De Wolf, I. M. Dremin, W. Kittel, Phys. Rept. **270**, 1 (1996).
- [65] PHENIX collaboration, A. Adare et. al., Phys. Rev. C **78** (2008) 044902.
- [66] T. Lappi, S. Srednyak, R. Venugopalan, JHEP **1001** 066 (2010).
- [67] CMS Collaboration, arXiv:1009.4122.
- [68] A. Dumitru, K. Dusling, F. Gelis, J. Jalilian-Marian, T. Lappi, R. Venugopalan, arXiv:1009.5295 [hep-ph].
- [69] A. Krasnitz, R. Venugopalan, Nucl. Phys. B **557**, 237 (1999); A. Krasnitz, Y. Nara, R. Venugopalan, Phys. Rev. Lett. **87**, 192302 (2001).
- [70] T. Lappi, Phys. Rev. C **67**, 054903 (2003); T. Lappi, Eur. Phys. J. C **55**, 285 (2008).